Title

Uncovering Privacy Leakage in BLE Network Traffic ofWearable Fitness Trackers

Authors

Das, Aveek K
Pathak, Parth H.
Chuah, Chen-Nee
et al.

Publication Date

2016-02-01

Peer reviewed

Uncovering Privacy Leakage in BLE Network Traffic of Wearable Fitness Trackers

Aveek K. Das, Parth H. Pathak, Chen-Nee Chuah, Prasant Mohapatra
University of California, Davis, CA, USA.
Email: {akdas, phpathak, chuah, pmohapatra}@ucdavis.edu

ABSTRACT

There has been a tremendous increase in popularity and adoption of wearable fitness trackers. These fitness trackers predominantly use Bluetooth Low Energy (BLE) for communicating and syncing the data with user's smartphone. This paper presents a measurement-driven study of possible privacy leakage from BLE communication between the fitness tracker and the smartphone. Using real BLE traffic traces collected in the wild and in controlled experiments, we show that majority of the fitness trackers use unchanged BLE address while advertising, making it feasible to track them. The BLE traffic of the fitness trackers is found to be correlated with the intensity of user's activity, making it possible for an eavesdropper to determine user's current activity (walking, sitting, idle or running) through BLE traffic analysis. Furthermore, we also demonstrate that the BLE traffic can represent user's gait which is known to be distinct from user to user. This makes it possible to identify a person (from a small group of users) based on the BLE traffic of her fitness tracker. As BLE-based wearable fitness trackers become widely adopted, our aim is to identify important privacy implications of their usage and discuss prevention strategies.

1. INTRODUCTION

The number of wearable devices shipped worldwide has had a growth of 200% from 2014 to 2015 [1]. Fitness trackers are by far the most popular wearable devices due to ever-increasing interest in the notion of *quantified-self* where users are able to track their daily activities (e.g. walking, physical workout, vital signs) with very high accuracy. The fitness trackers connect to user's smartphone using a short-range wireless communication like Bluetooth Low Energy (BLE). Due to substantial reduction in energy consumption, BLE has become the dominant standard for the fitness trackers to connect and communicate with smartphones.

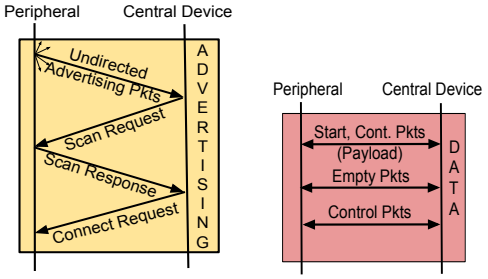
Although the fitness trackers and BLE are becoming widely used, the private information that leaks through the BLE communication has largely remained unexplored. In a recent study [2], it is shown that motion sensors on wrist-worn devices (like fitness trackers) can leak the information about what a user is typing. Different from this, we explore how private information about the user can be leaked by eavesdropping on BLE communication between the fitness trackers and the smartphone. As BLE becomes pervasive with its adoption for Internet of Things and proximity sensing services like iBeacons in public places, eavesdropping BLE communication can be easier than ever before, making it imperative to protect user's privacy. In this paper, we present a measurement study by collecting BLE traffic between fitness trackers and smartphones, and discover the following privacy leakage -

(1) User Tracking: We show that almost all fitness tracker devices utilize unchanged BLE addresses, making the user vulnerable to tracking. Specifically, fitness tracker and smartphone only periodically connect to each other for exchanging data, leaving the fitness tracker in disconnected advertising mode most of the time where it constantly announces its presence by broadcasting advertising packets. This continuous advertising by the fitness trackers using unchanged BLE addresses can be combined with additional information (e.g. video monitoring) by an attacker to track the owner of the BLE device. Using traces collected in a gymnasium as well as in controlled experiments, we find that the issue prevails in over 90% of observed devices including top five leading fitness tracker manufacturers, namely Fitbit, Jawbone, Polar, Garmin and Misfit. BLE standard [3] outlines the use of randomized addresses for prevent tracking, however, it is optional and we find that majority of the fitness tracker manufacturers do not follow them in practice. Compared to tracking through WiFi MAC address (recently addressed in [4]), BLE tracking can provide more fine-grained location of user due to its smaller range and is also feasible even when user's smartphone is connected to a cellular network.

(2) User Activity Detection and Person Identification: We find that the BLE data traffic between a fitness tracker and a smartphone is correlated to the intensity of user's activity. This means that simply by observing and analyzing the BLE traffic, an eavesdropper can detect user's current activity such as walking, sitting, running etc. For example, a employer can track the activities of employees by deploying sniffers in the office space. Our evaluation shows that the activity recognition is feasible with 97.6% for 10 users.

Furthermore, we show that there is a strong correlation between the motion sensor (accelerometer) readings of wrist-worn fitness tracker and the patterns of its BLE traffic to the smartphone. Based on the fact that different users walk with distinct gait, an eavesdropper can analyze the BLE traffic and uniquely identify the user from a small group of users. This means that a fitness tracker user can be identified through BLE traffic analysis even when the fitness tracker randomizes its BLE address. We derive the necessary BLE traffic attributes and show that person identification is feasible with an accuracy of 89% for groups of 5 people. Compared to address-based device tracking where a person can be tracked anywhere, identification of a person is only possible from a small group of fixed

HotMobile '16, February 26-27, 2016, St. Augustine, FL, USA
 ISBN 978-1-4503-4145-5/16/02...$15.00
DOI: http://dx.doi.org/10.1145/2873587.2873594



(a) Advertising Packets (b) Data Packets

Figure 1: BLE Communication - Advertising & Data

users, making it a privacy risk in cases like homes, offices etc.

2. BLE OVERVIEW AND SNIFFING

In this section, we provide an overview of BLE communication. We focus on the aspects necessary in understanding the BLE privacy leakage. We also discuss the challenges related to sniffing BLE traffic and our data collection methodology.

2.1 BLE Background

2.1.1 BLE Communication

BLE operates in the 2.4 GHz ISM band and utilizes 40 RF channels with 2 MHz spacing. The BLE communication is divided in two phases - advertising and data communication.

Advertising: This phase is responsible for device advertisement, device discovery and establishing a connection. Packets in the advertising phase are sent out in the 3 dedicated channels. A BLE peripheral device (e.g. a fitness tracker) transmits advertisement packets to announce its presence to the master device (e.g. a smartphone). In BLE communication, each packet is associated with an access address which uniquely identifies a connection between two devices. The flow diagram of the advertising packets is shown in Fig. 1a. As we will discuss later, fitness tracking devices commonly use *Undirected Connectable Advertising Packets* in order to allow any master device to connect to it. The advertising packets contain information like MAC address, connectability modes and TX(transmission) power level. A master device upon receiving an advertising packet, if interested in initiating a connection, can send a *Scan Request* to the peripheral requesting additional information such as device local name, supported profiles, etc. The peripheral responds with a *Scan Response* which contains these additional information not included in the initial advertising packet. The master device then establishes a connection using a *Connect Request* along with further exchange of information (e.g. sharing of keys for secure connection). Other types of advertising packets are *Connectable Directed, Non-Connectable Undirected and Scannable Undirected* - used by devices to establish a quick connection or by devices which act just as transmitters.

Data Communication: Once a BLE peripheral is connected to a master device, the communication is carried out over the 37 data channels using adaptive frequency hopping. In the data communication phase, a new access address is used every time the master and peripheral reconnect. Most of the data communication (transfer of data payload) in BLE happens through the use of *Start* and *Continuation* packets. When new data is being exchanged between the devices, a start packet is used which is then followed by one or more continuation packets if more bytes are needed to be transferred. When two devices are connected, meaningful data is normally sent out in bursts (in order to save energy). On the other hand, the devices hop from one frequency to another in very short intervals. During each hop, if no meaningful data is to be transmitted, empty packets are exchanged before the devices hop to a new channel. These packets have no payload and just consist of packet headers. In addition, we also see *Control* packets which are used for updating of connection parameters (like hop interval, access address, etc.) and for connection termination.

2.1.2 Private advertising addresses in BLE

Compared to Bluetooth classic, BLE introduces the use of random addresses, whereby the real address (i.e. MAC address) of a BLE device is hidden and a random address (which changes frequently) is advertised. BLE devices can use manufacturer provided fixed MAC address as its address or optionally choose one of three types of random addressed described below [3].

1) Static address: A BLE device uses a randomly generated address that either changes only at bootup or always remains unchanged. This type provides the least privacy against device tracking, especially if the address remains unchanged.

2) Non-resolvable Private address: The address changes periodically and provides better privacy compared to the static addresses.

3) Resolvable Private address: It is generated using a Identity Resolving Key (IRK) and a random number. The advantage of this type of address over non-resolvable is that it can be resolved using the shared IRK to uniquely identify a device.

The type of random address can be detected by looking at two Most Significant Bits (MSB) of an address (11 - static address, 00 - non-resolvable private address and 10 - resolvable private address).

2.2 BLE Network Traffic Sniffing

There are two main challenges in sniffing BLE traffic. First, when a BLE peripheral is in advertising phase, the advertising packets are transmitted on all three advertising channels by periodically switching between them. Since the connection can be established on any of the three channels, it is necessary to sniff all three advertising channels in parallel. Second, once the connection is established, the sniffer should be able to follow the hopping sequence (channel map) of the connection to sniff each data packet on 37 data channels. We use ComProbe Bluetooth Protocol Analyzer (BPA) 600 [5] for sniffing. It can capture BLE advertising packets on all 3 channels and can follow a connection over data channels after the connection is established. The analyzer software (shown in Fig. 2a) allows us to investigate each filed of BLE packets. We note that popular open-source BLE sniffing platform - Ubertooth [6] - can also be used, however, it is limited to sniffing only one advertising channel at a time and provides very few dissectors for traffic analysis.

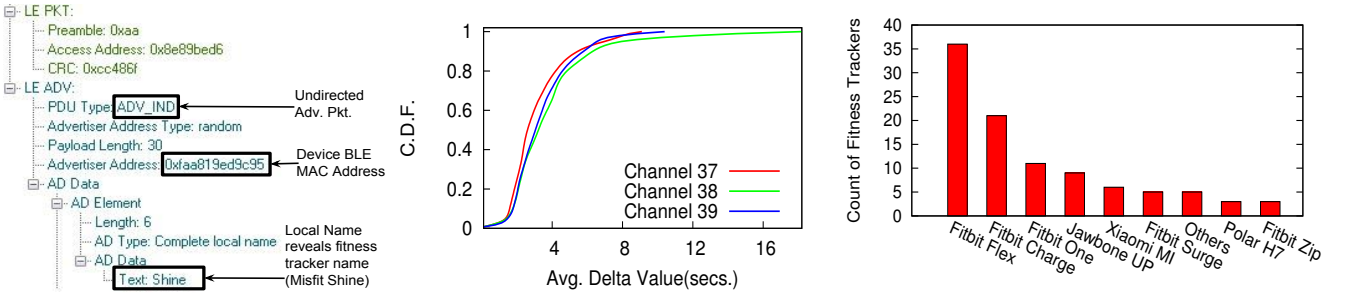
3. DEVICE TRACKING USING ADVERTISING PACKETS

In this section, we investigate the private information leaked about the user from the advertising packets of her fitness tracker. We study how the information leakage can lead to tracking of the user.

3.1 BLE Dataset

In order to study the privacy leakage from advertising packets, we collect network traffic traces in the form of two datasets -

Gym Dataset: For understanding BLE traffic in the wild, we collect traffic traces in a gymnasium where there are likely to be more users with fitness trackers. We collect the network traffic traces by sniffing the packets in the air using the ComProbe BPA 600. We only capture the packets for BLE (and not Bluetooth Classic) as most of the fitness trackers use BLE for communicating with smartphone. We primarily focus on collecting advertising packets



(a) Snapshot of the information in an Advertising Packet

(b) C.D.F of the average delta values for the fitness trackers in the three advertising channels

(c) Different types of fitness trackers - revealed by the field *Local Name* in advertising packets. "Other" includes Garmin, Misfit, Samsung and Empatica.

Figure 2: BLE Dataset: Snapshot of Packet in Data Collected, Packet intervals distributions , Fitness Trackers in the Dataset

in this data collection. The traces were collected for 8 consecutive days, with each trace being two hours in duration. Overall, the dataset contained a total of around 7.5 million packets with the total size of traces being 3.5 GB. Table 1 shows the number of advertising packets and devices for each day, and average packets per second. Fig. 2b represents how frequently the advertising packets are sent out by showing the C.D.F. of the mean value of time interval between two consecutive packets in the same channel for each device. We note here, that in each interval calculated, there are two more packets being sent out in the other two advertising channels. As per BLE standard, the duration between two advertising packets should vary between 20 ms and 10.24 seconds. We observe that most values of the this interval (about 95%) are less than 8 seconds, proving that these devices transmit advertising packets continuously.

Apart from the BLE MAC addresses, which we anonymize, the collected BLE traffic contains no user-identifiable information (such as user names or email addresses). An attacker can associate a MAC address to a specific user over a period of time with the use of auxiliary information about user's presence (such as video recording). We do not acquire an IRB approval because we do not collect any such additional information in our dataset and only rely on passively monitoring network packets without direct user involvement.

Controlled Dataset: For controlled experiments, we use 6 popular fitness trackers (Fitbit Flex, Fitbit Charge, Jawbone UP, Garmin vivosmart, Misfit Shine and Polar Heart Rate Sensor) and collect its BLE traffic over multiple days in a controlled setup using the ComProbe BPA 600 when these fitness trackers are connected to an iPhone 6 and a Nexus 6.

We empirically evaluate the range of the sniffer to be approximately between 20 to 25 meters. Fig. 2a shows a sample advertising packet and its fields such as device's advertised address, access address (fixed for advertising) and "Complete Local Name".

3.2 Consistent Advertising and Static BLE Addresses

For the gym dataset, we first determine if the advertising BLE device is a fitness tracker using the "Shortened Local Name" or "Complete Local Name" field. We observe that the local name field reveals the device name in plain-text. For example, a fitness tracker "Fitbit Flex" has the "Shortened Local Name" of *Flex* in the advertising packet. In Fig. 2a we see the complete local name *Shine* as a part of the packet - which indicates that the device is a "Misfit Shine". Using this information, we determine that there are 99 distinct fitness trackers in the collected traces. Fig. 2c shows the manufacturer and model of the fitness trackers in our dataset as determined from the complete local name. Note that there can be many other fitness trackers that do not reveal their name in the advertising packets. The other types of BLE devices observed in the traces, that

Trace	Advertising Packets (millions)	Advertising Devices	Fitness Trackers	Avg. Pkts. per sec.
A	0.504	189	12	69.1
B	.41	147	7	60.2
C	0.76	200	12	99.7
D	1.07	182	18	145.2
E	1.40	207	24	196.8
F	1.28	226	21	172.6
G	1.09	277	21	125.7
H	.99	188	12	147.3

Table 1: BLE Packet Traces collected from Gymnasium

are not fitness trackers, primarily include gym equipment like body scales and treadmills. In our traces, we do not find any devices which we could identify as a smartwatch. We also observe that the smartphones (both iOS and Android), being master devices, do not advertise continuously and also change their addresses (which is observed in *Connect Request*). The only devices, apart from fitness trackers, which have a "Local Name" in our dataset are gym equipment.

Why fitness trackers constantly advertise? The high frequency of advertising packets by almost all fitness trackers raise an important question - why the fitness trackers consistently advertise even when they are in close proximity of owner user's smartphone? Through the controlled dataset, we observe that the frequent advertising is due to the fact that the master device (i.e. smartphone) frequently disconnect the fitness trackers in order to reduce its own energy consumption. The fitness trackers are only connected to the smartphone when the corresponding smartphone application on the smartphone is running (foreground). When the app is running, the tracker actively communicates and synchronizes the activity data (e.g. steps, calories etc.). When the app is not running, the connection is terminated, leaving the fitness tracker in advertising state. This behavior was observed in all six fitness trackers in the controlled dataset.

The constant advertising of BLE fitness trackers make them vulnerable to tracking. This means that an attacker can sniff the BLE traffic and track the users' visits as they move around in public places such as shopping malls, gymnasiums, cafeterias etc. As discussed in Section 2.1.2, the BLE devices can choose to change their address in order to avoid tracking. However, as we discuss next, majority of the BLE fitness trackers do not use the random addresses, leading to a severe privacy implication of user tracking through fitness trackers.

Unchanged BLE Addresses: Through our controlled experiments, we observe that none of the six fitness trackers change their BLE address. We power-cycle the devices multiple times by draining their battery and find that the addresses do not change after reboot. We also observe that the advertised address do not have the

Trace	A	B	C	D	E	F	G	H
A	-	1	2	0	2	0	0	1
B	1	-	1	1	0	0	0	0
C	2	1	-	0	1	0	1	1
D	0	1	0	-	1	1	2	1
E	2	0	1	1	-	**4**	2	0
F	0	0	0	1	**4**	-	**4**	2
G	0	0	1	2	2	**4**	-	**4**
H	1	0	1	1	0	2	**4**	-
Total	5	3	4	6	8	9	10	7

(a) Reappearing Fitness Trackers

Trace	A	B	C	D	E	F	G	H
A	-	3	9	2	**16**	1	3	4
B	3	-	5	4	1	**16**	1	2
C	9	5	-	0	3	2	3	2
D	2	4	0	-	5	4	**21**	2
E	**16**	1	3	5	-	6	8	4
F	1	**16**	2	4	6	-	**13**	8
G	3	1	3	**21**	8	**13**	-	9
H	4	2	2	2	4	8	9	-
Total	30	27	15	32	34	40	45	21

(b) Reappearing BLE Devices

Table 2: Reappearing Device Count Matrix - Each element of the matrix is the number of BLE devices that are common between two traces.

Address Type	MSB	% Fitness Trackers
Non-resolvable Private Address	00	0
Resolvable Private Address	10	11
Static Address	11	89

Table 3: Fitness Trackers and their different address types

MSB of 00 or 10 (Section 2.1.2) - the recommended standard for private addresses. Thus, we conclude that these devices never alter their address. With additional information which can map a specific device to a specific user, the unchanged BLE addresses make the users carrying these devices trackable.

To further confirm our observation about unchanged device addresses in the controlled experiments, we analyze the gym dataset. Table 3 shows the MSB of the observed fitness trackers in our dataset where we see that majority (89%) of them use static address with MSB of 11. Furthermore, Table 2 shows a matrix where each element is the number of common devices (same advertising address) across the two traces. We show the matrix for all BLE devices and the ones which we know are fitness trackers based on their advertised names. We find that -

(1) 113 devices out of a total of 1485 BLE devices have appeared in more than one trace.

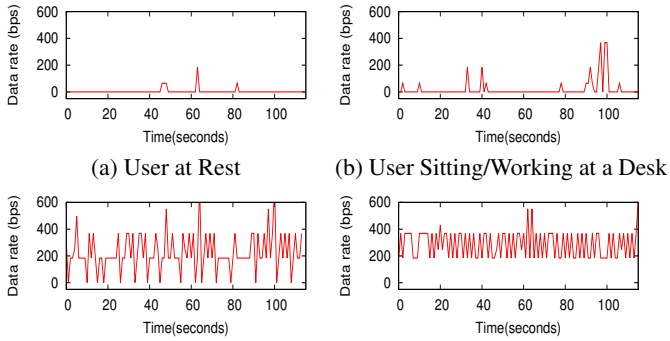
(2) 24 out of the total 99 fitness trackers have appeared in more than one trace. This is number is noticeably high (almost 25%) given that the traces were only collected on different days for only two hours per day.

(3) The highest overlap in advertising devices between two traces is 21, whereas the highest for fitness trackers is 4.

The number of reappearing devices proves our hypothesis inferred from the controlled experiments - that the BLE devices do not alter their advertising addresses - making its users vulnerable to tracking through the use of auxiliary information which can map a specific device to its user.

4. ACTIVITY AND PERSON IDENTIFICATION USING DATA PACKETS

As described in Section 2, once the advertising device (fitness tracker) receives a connection request from the master device (smartphone), the data transfer phase starts and data is transferred on all 37 channels. BLE utilizes AES-CCM encryption method and it is difficult to decrypt the payload without intercepting packets in initial key exchange phase [7]. In this section, we show how statistical traffic pattern analysis over the encrypted traffic can be used to detect user's activity. In scenarios, when the BLE devices actually use the private address techniques and alter their advertising addresses regularly, this method can also be used to identify an individual from a small user group. Since Fitbit fitness trackers are by far the most widely adopted devices (from [1] and Fig. 2c), we only focus on Fitbit devices (specifically, Flex) in this section.



(a) User at Rest (b) User Sitting/Working at a Desk

(c) User Walking (d) User Running

Figure 3: Data Rate for different User Activities

Experimental Setup: We conduct experiments with 10 volunteers who wear a Fitbit on their non-dominant wrist and walk in their natural gait. We also attach a smartphone (Nexus 6) close to the Fitbit on user's wrist to collect raw accelerometer signals at 20Hz sampling rate. The BLE traffic between the Fitbit and user's smartphone is sniffed using ComProbe BPA 600. Each experiment lasts for 150 seconds and is repeated 10 times per user. The data is collected while the Fitbit app is open in the foreground on an iPhone 6. In addition to the walking experiments, we collect the accelerometer and BLE traffic data while the user is at rest, working at a desk and running.

4.1 Activity Detection

When the smartphone application of Fitbit is open and running in foreground, the Fitbit and the smartphone actively exchange data packets. The data communication stops when the smartphone application is closed. We observe that the amount of BLE data traffic between the Fitbit and the smartphone is proportional to the (motion) intensity of user's activity. For example, when the user is sitting with some sporadic low-intensity motion (desk-bound or sedentary), the BLE traffic consists of a large number of empty packets and only a small number of start or continuation packets. This means that the size of data being transferred is much relatively smaller in size. In comparison, when the user is walking, the data to be transferred increases, resulting in more number of start packets. We observe that even though the packets are encrypted, the volume of the data is a definite indicator of the user's activity. The privacy implication is that an attacker can sniff the BLE traffic and infer user's current activity. For example, an employer can track and monitor the activities (e.g. sitting on desk, walking etc.) of employees at workplace, and if the fitness trackers also don't change their device address, the employer can even track employee's walking trajectories using multiple sniffers. In a gymnasium, an attacker can monitor the amount of time a user sits, walks or runs everyday.

In BLE data communication, meaningful data is transmitted in *Start* and *Continuation* packets. We refer to the data transmitted in these packets (without their header) as the BLE payload. Fig. 3

Accelerometer	Mean and Max Acceleration Zero-Crossings, Absolute Area Sum of Absolute Acceleration
BLE Data	Start Pkts., Empty Pkts., Payload Size Payload Datarate, Time b/w Start Pkts. Empty Pkts. b/w Start Pkts

Table 4: Feature Set : Accelerometer Features for Correlation, each measured on each of the 3 axis (X,Y,Z). BLE Data features for person identification. We calculate min, max, mean and standard deviation for the last two BLE Data features.

shows the payload data-rate in bits per second for four different user activities, namely, stationary (at rest or sleeping), sitting (or working at a desk), walking and running. We observe that the data rate of the Bluetooth communication is different from when the user is at rest or in motion. A comparison of Figs. 3a and 3b shows that when a user is working, the data rate for most parts is similar to a stationary user apart from when the user moves her hands which results in spikes in data transmission, as seen in the later part of the Fig. 3b. Since, Fitbit is not just a step-counting device, we note here that these spikes are not necessarily due to a step being reported, but also due to the update of other information (like calories). Comparing the data rates of walking and running, we observe that the data rate does not fall to *zero* for the running case Fig. 3d, confirming the proportionality with the intensity of activity.

We further validate the claims of activity detection through BLE traffic analysis using the data collected from 10 volunteers. Using the collected data we calculate a feature vector for time windows of 20 seconds. The feature vector includes (1) payload data rate, (2) number of empty packets and (3) number of start packets. Using the feature vector, a decision tree classifier can classify the 4 activities with an accuracy of 97.6%.

4.2 Person Identification

Wearable fitness trackers calculate a number of useful health-related statistics like number of steps walked, total calories burnt, total distance covered, flights of stairs climbed etc. When the smartphone app is running in the foreground, these information is sent from the Fitbit to the smartphone, which in-turn updates the user interface on the app. In the previous subsection, we observed that the intensity of a user activity is related to the data-rate of the Bluetooth connection. In this section, we show that the BLE data exhibits different patterns when different users are walking, making it possible to uniquely identify a user from a small group of users.

Correlation with Accelerometer Data: Fitbit Flex utilizes a 3-axis accelerometer to monitor user movements (frequency, duration and patterns) and derive necessary statistics [8]. A later model of Fitbit (Surge) also has a gyroscope, compass and ambient light sensor, but for our experiments we just focus on the accelerometer readings. Since the actual algorithms used by Fitbit to monitor user activities are unknown, we conjecture that there is a strong correlation between the observed accelerometer signal and corresponding BLE traffic. If this correlation is indeed strong, the BLE traffic can be used by an attacker to detect user's walking speed and gait. As we know from past research [9] that user's gait can uniquely identify the user with high accuracy (especially in a small group of users), the BLE traffic can also be misused for user identification.

Using the collected acclerometer and BLE data for walking activity of 10 volunteers, we calculate the statistical features listed in Table 4 for 20 seconds time windows. The accelerometer features we use are found to be useful in detecting human physical activities in [10]. We then build separate linear regression models which use these accelerometer features as input to predict each of the BLE network features. We compute the correlation coefficient between the calculated values of the BLE network features (using regression) and the actual values obtained from the captured data. The correlation coefficients for different BLE traffic features are listed in Table 5. We observe a correlation of approximately 70% for payload datarate and empty packet count. This shows that the BLE traffic is correlated to the observed accelerometer data.

BLE Data Feature	Correlation
Empty Pkts.	0.705
Payload Datarate	0.699
Start Pkts.	0.684
Payload Size	0.676
Time b/w Start Pkts.	0.647
Pkts b/w Start Pkts.	0.634

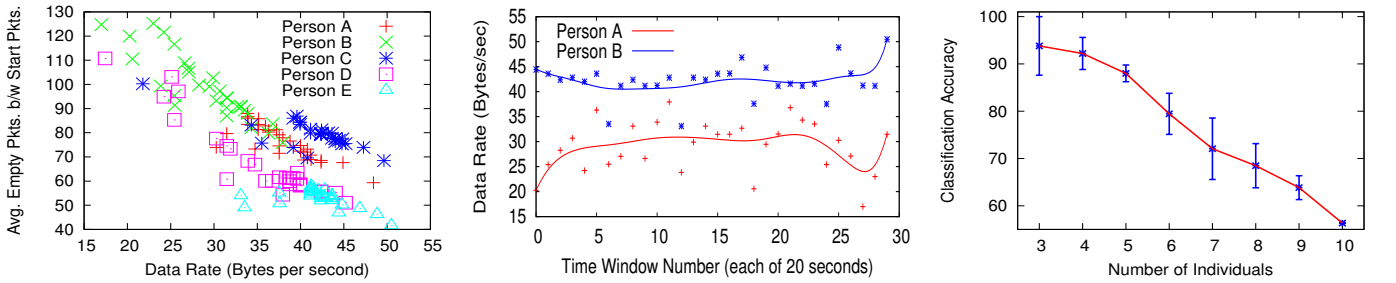
Table 5: Correlation between predicted values of BLE Data feature, calculated using linear regression on accelerometer values, and the actual value of the feature.

Person Identification using BLE Traffic: Because of the correlation of BLE traffic pattern with the accelerometer data, it represents user's gait while walking and thereby can be used to uniquely identify the user. Based on the BLE network data collected for 10 users, we calculate the features shown in Table 4 for 20 second windows. Fig. 4a shows two features - payload data rate and average number of empty packets between two start packets - for 5 users. We observe that for each user there is a non-overlapping cluster signifying the two BLE features can distinguish the 5 users. We also represent the BLE payload data rate for two representative users in 4b and see the variation is very distinct. Thus, the features extracted can be considered useful for uniquely identifying a user. We use the BLE features and build a person identification classifier using decision tree. Fig. 4c show the average accuracy of person identification using the BLE traffic features. We consider all possible combinations of users when the person identification classifier is built for less than 10 users. The standard deviation of accuracy for different user sets are also shown in Fig. 4c. The false positive rate for all the different user sets was less than 5%. The classification accuracy decreases with increase in number of users because dissimilarities in gait reduces as user population increases.

Person identification through BLE traffic analysis is a major privacy concern as it can enable an attacker to track a user using her fitness tracker even when it changes its BLE address. The person identification works when a classifier is pre-trained for the walking pattern of a known set of users. In many cases such as office buildings or gymnasiums where the same set of users reappear frequently, the collected data can be used to train an accurate model. It is to be noted here that the attack model we discuss works when the Fitbit app is open in the foreground. However in the case of certain trackers, like Garmin, the attack is possible even when the app is not running in the foreground, as long as the phone and the BLE device has been previously paired.

5. RELATED WORK

In recent years, there has been a number of research works on the leakage user's activity through wireless signal analysis. Keystroke recognition [11] and human activity recognition [12] has become possible through changes in the Channel State Information(CSI) as a user moves or types on her keyboard. [13] shows how the accelerometer and gyroscope sensors in smartwatches can be used to uniquely identify finger movements, hand and forearm motion of users on the basis of some essential features extracted from these sensors' data. This paper also shows that finger-writing on a surface or on the air can be detected from these sensors, whereas [2] shows that when a user is typing on the keyboard the motion sensors on a smartwatch can predict the word typed out with a certain level of confidence. Our work differs from these, as we focus on device tracking and user activity detection just from the point of



(a) Data Rate vs Mean of Empty Packet count between Start Packets for 5 different users

(b) BLE Data Rate variation for two representative users among the 10 volunteers.

(c) Mean Accuracy ± Standard Deviation with different number of Individuals

Figure 4: BLE Person Identification Results

view of the network data created by the devices (and not based on the sensor values themselves).

In mobile telephony, temporary and contextual user identifiers have been proposed instead of the permanent ones to prevent tracking but some of these techniques have not been very successful in providing privacy [14]. From the point of view of Bluetooth traffic, a Man-in-the-Middle attack on a Bluetooth keyboard allows an attacker to access all keystrokes on a keyboard and lead to serious security leaks. Similar attacks has been executed using data from wireless mouse to reconstruct mouse cursor trajectory and infer private user information [15]. Some researchers proposed the use of the device clocks (time interval among advertising packets) to fingerprint different Bluetooth devices and prevent address forging. There has been recent some research which looks at how BLE data communication security can be broken by capturing the necessary keys during the connection-establishment phase [7]. Compared to this, our work does not attempt to decrypt the encrypted BLE traffic but focuses entirely on feasibility of mining already encrypted BLE data from the point of view of activity and person identification.

6. DISCUSSION AND CONCLUSION

The advertising and data communication phases of BLE network traffic cause concerns from the point of view of user privacy. The detection of activity is possible due to the fact that sending of BLE data (payload) is triggered only by user activity. This can be prevented by sending out artificial traffic (or chaff) [16]. In this solution, we can insert artificial data packets (start and continuation) in our BLE network so that activity recognition from the payload pattern becomes more complex and circumvent traffic analysis. However, one drawback of this is that the energy consumption would increase as a result of transmitting more packets than required. Thus, a balance has to be maintained between sending out artificial traffic at certain intervals so as to prevent detection but not at the expense of high energy consumption.

To prevent user tracking based on advertising packets one potential solution is to randomize the advertised address, a topic that has recently been brought to light for WiFi communication, with the recent versions of iOS randomizing the MAC address while broadcasting to prevent tracking at public places. There has also been efforts in terms of mobile telephony to use temporary identifiers instead of long term permanent identifiers to prevent third-party tracking. However, usage of randomized addresses can lead to the smartphone not being able to identify the BLE fitness tracker to which is has already been paired and might need the tracker to pair again - leading to a disruption in user experience. Also, randomized addresses can still be used to track a user based on the user activity determined using data packets. Another solution is not to advertise continuously and instead, use direct advertising packets from the fitness trackers directly to the user's smartphone (to which the tracker has been previously synced), when the smartphone switches on the fitness tracker app. With the ever increasing popularity of smartwatches and corresponding applications on these devices, BLE communication privacy is even more critical in the near future. In our future work, we will analyze BLE network data generated by different applications in smartwatches from the point of view of user privacy.

7. REFERENCES

[1] "IDC Worldwide Quarterly Wearable Tracker, June 2, 2015." http://www.idc.com/getdoc.jsp?containerId=prUS25658315.

[2] H. Wang, T. T.-T. Lai, and R. Roy Choudhury, "Mole: Motion leaks through smartwatch sensors," ACM MobiCom '15.

[3] "Bluetooth Low Energy 4.1 Standard," *Bluetooth SIG., Inc. Specification of the Bluetooth System*, 2013.

[4] "Randomized Wi-Fi addresses." http://www.apple.com/lae/privacy/privacy-built-in/.

[5] "ComProbe BPA 600." http://www.fte.com/products/BPA600.aspx.

[6] "Project Ubertooth." http://ubertooth.sourceforge.net/.

[7] M. Ryan, "Bluetooth: With low energy comes low security," in *7th USENIX Workshop on Offensive Technologies*, 2013.

[8] "Fitbit Help: How does my tracker count steps?." http://help.fitbit.com/articles/en_US/Help_article/How-does-my-tracker-count-steps.

[9] L. Rong, D. Zhiguo, Z. Jianzhong, and L. Ming, "Identification of individual walking patterns using gait acceleration," in *ICBBE 2007*.

[10] E. Munguia Tapia, *Using machine learning for real-time activity recognition and estimation of energy expenditure*. PhD thesis, MIT, 2008.

[11] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using wifi signals," ACM MobiCom '15.

[12] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," ACM MobiCom '15.

[13] C. Xu, P. H. Pathak, and P. Mohapatra, "Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch," ACM HotMobile '15.

[14] M. Arapinis, L. Mancini, E. Ritter, and M. Ryan, "Privacy through pseudonymity in mobile telephony systems," NDSS '14.

[15] X. Pan, Z. Ling, A. Pingley, W. Yu, N. Zhang, and X. Fu, "How privacy leaks from bluetooth mouse?," ACM CCS '12.

[16] S. Le Blond, D. Choffnes, W. Zhou, P. Druschel, H. Ballani, and P. Francis, "Towards efficient traffic-analysis resistant anonymity networks," ACM SIGCOMM '13.